JOURNAL OF BACTERIOLOGY, Oct. 1988, p. 4582–4588
0021-9193/88/104582-07$02.00/0

Vol. 170, No. 10

Nucleotide Sequence and Deletion Analysis of the Xylanase Gene (*xynZ*) of *Clostridium thermocellum*

OLIVIER GRÉPINET, MARIE-CHRISTINE CHEBROU, AND PIERRE BÉGUIN*

Unité de Physiologie Cellulaire, Département des Biotechnologies, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France

Received 7 March 1988/Accepted 5 July 1988

The nucleotide sequence of the *xynZ* gene, encoding the extracellular xylanase Z of *Clostridium thermocellum*, was determined. The putative *xynZ* gene was 2,511 base pairs long and encoded a polypeptide of 837 amino acids. A region of 60 amino acids containing a duplicated segment of 24 amino acids was found between residues 429 and 488 of xylanase Z. This region was strongly similar to the conserved domain found at the carboxy-terminal ends of *C. thermocellum* endoglucanases A, B, and D. Deletions removing up to 508 codons from the 5′ end of the gene did not affect the activity of the encoded polypeptide, showing that the active site was located in the C-terminal half of the protein and that the conserved region was not involved in catalysis. Expression of xylanase activity in *Escherichia coli* was increased up to 220-fold by fusing fragments containing the 3′ end of the gene with the start of *lacZ* present in pUC19. An internal translational initiation site which was efficiently recognized in *E. coli* was tentatively identified 470 codons downstream from the actual start codon.

The gram-positive anaerobic thermophile *Clostridium thermocellum* secretes a highly active and thermostable cellulolytic complex (15, 18, 28). Most natural cellulose is closely associated with hemicellulose. Xylans are major components of hemicellulose. They are heteroglycans with a backbone of (1,4)-linked β-D-xylopyranosyl residues and short side chains of (1,3)-linked α-L-arabinofuranose and (1,2)-linked α-D-glucopyranuronic acid residues (3).

C. thermocellum secretes xylanase activity when cells are grown on cellobiose (9), but growth on xylan occurs only after a lag phase of several days (46). *C. thermocellum* is able to utilize xylooligomers of $n = 2$ to 5 but not xylose, which accumulates in the culture medium during growth on xylan (46). Xylanolytic enzymes secreted by *C. thermocellum* during growth on cellobiose may make cellulose accessible to cellulolytic enzymes. No xylanase from *C. thermocellum* has yet been isolated and characterized.

Ten clones containing distinct chromosomal DNA fragments from *C. thermocellum* were isolated by Millet et al. (24) and were tentatively identified as carrying cellulase (*cel*) genes on the basis of their ability to hydrolyze carboxymethylcellulose or 4-methylumbelliferyl-β-D-cellobioside (MUC). However, further analysis showed that the activity towards MUC of clones carrying pCT1200 and pCT1300 was due to xylanase rather than cellobiohydrolase activity (data not shown).

Plasmid pCT1200 confers on *Escherichia coli* the ability to hydrolyze xylan, MUC, para-nitrophenyl (pNP)-β-D-cellobioside, pNP-β-D-xylopyranoside, and pNP-β-D-xylobioside but not carboxymethyl cellulose. We have shown that a single protein termed xylanase Z, encoded by the gene *xynZ*, is responsible for these activities (accompanying paper [11]). We report here the structure of the *xynZ* gene, the similarities found with endoglucanases from *C. thermocellum*, and the construction of deletions in the *xynZ* gene.

* Corresponding author.

MATERIALS AND METHODS

Enzymes and reagents. All restriction endonucleases were purchased from Amersham or Boehringer-Mannheim and were used as recommended by the suppliers. T4 DNA ligase and exonuclease III (Boehringer-Mannheim), the Klenow fragment of DNA polymerase and S1 nuclease (Amersham), and mung bean nuclease (New England Biolabs) were used as recommended by the suppliers. Isopropyl-β-D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside were from Sigma.

Bacterial strains and vectors. Phages M13mp8 and mp9 (23), cloning vectors pUC8 (42) and pUC19 (48), and plasmid pTZ18R (Pharmacia Ltd.) were prepared in *Escherichia coli* TG1 (43). pCT1200 is a derivative of pACYC184 (5) which contains a 6.0-kilobase (kb) *Eco*RI fragment (Fig. 1) carrying the *xynZ* gene of *C. thermocellum* (24).

DNA isolation and fractionation. Plasmid DNA and RF M13 DNA for restriction analysis were isolated by the alkaline lysis procedure (4). Large-scale plasmid purification was obtained by banding in CsCl-ethidium bromide density gradients (14). DNA restriction fragments were resolved by horizontal gel electrophoresis in borate buffer (21).

Nucleotide sequence analysis. Overlapping deletions were generated with exonuclease III as described by Guo and Wu (12), except that mung bean nuclease was used instead of S1 nuclease. Fragments of appropriate size were sequenced in M13mp8 and mp9 and pTZ18R by the dideoxy chain termination method of Sanger et al. (34). The entire coding sequence was determined at least once on both strands.

Construction of deletions of *xynZ*. The 6.0-kb *Eco*RI fragment of pCT1200 (Fig. 1) was cloned into pUC8 to yield pCT1202. The 2.15-kb *Acc*I-*Cla*I fragment was treated with the Klenow fragment of DNA polymerase and cloned at the *Sma*I site of pUC8 to yield pCT1208 (Fig. 1). Plasmid pCT1211 contains the same fragment as pCT1208 but cloned in pUC19. *xynZ* is transcribed in the same direction as *lacZ* in pCT1202, pCT1208, and pCT1211. Plasmid pCT1223 was constructed by removing the 50-base-pair (bp) *Eco*RI-*Sty*I fragment of pCT1208 and recircularizing the plasmid after

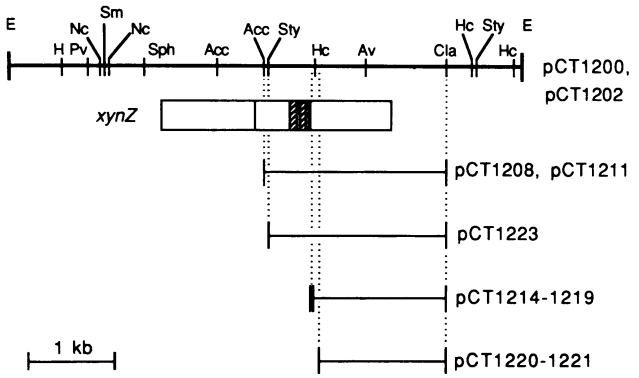


FIG. 1. Position of *xynZ* within the 6.0-kb *Eco*RI fragment carried by pCT1200. The coding sequence is shown by a rectangle. The protein is translated from left to right. In fragments cloned in pUC vectors, *xynZ* has the same orientation as *lacZ*. The duplicated segment is hatched. Pro- and Thr-rich regions are shown by solid bars. Inserts carried by various deleted subclones are shown by horizontal bars. The exact endpoint of each deletion is shown in Fig. 2. Restriction sites: Acc, *Acc*I; Av, *Ava*I; Cla, *Cla*I; E, *Eco*RI; H, *Hin*dIII; Hc, *Hin*cII; Nc, *Nco*I; Pv, *Pvu*II; Sm, *Sma*I; Sty, *Sty*I.

trimming single-stranded ends with mung bean nuclease. Plasmids pCT1214 to pCT1221 were constructed by adapting the exonuclease III method of Guo and Wu (12). Plasmid pCT1211 was cut by *Pst*I and *Sty*I, digested with exonuclease III and mung bean nuclease successively, and then ligated with itself. The endpoint of these deletions was determined by double-stranded DNA sequencing (6) with the M13 reverse sequencing primer (Biolabs).

Specific xylanase activity of crude extracts. Clones were grown overnight in 200 ml of LB medium containing carbenicillin (100 μg/ml). Cells were centrifuged at 5,000 × *g* for 20 min, suspended in 10 ml of PC buffer (50 mM K_2HPO_4, 12.5 mM citric acid, pH 6.3), and disrupted by sonication in a Branson B-12 sonifier. The extract was centrifuged at 5,000 × *g* for 15 min to remove cell debris. Xylanase was assayed by incubating crude extracts in a 0.5% (wt/vol) solution of xylan (larchwood xylan, Sigma) in PC buffer at 60°C. The appearance of reducing sugars was assayed by the Somogyi-Nelson method (27). One unit of activity corresponds to the release of 1 μmol of xylose equivalent per min. Protein concentration was determined by the Coomassie blue G-250 binding assay (35) with bovine serum albumin as a standard.

Immunological detection of *xynZ* gene products. Crude extracts from various clones were analyzed by Western blotting (immunoblotting) (39). The anti-xylanase Z antiserum (11) was saturated with a crude extract of *E. coli* TG1(pUC19) to prevent adsorption to nonspecific bands.

RESULTS

Sequence of the *xynZ* gene. The region of pCT1202 expressing xylanase Z was found to contain an open reading frame of 2,511 nucleotides (Fig. 2). The encoded 837-amino-acid polypeptide had a calculated molecular weight of 92,159, in close agreement with the M_r of 90,000 found for xylanase Z by Western blotting both in crude extracts of *E. coli* TG1(pCT1202) (see Fig. 4, lane 2) and in *C. thermocellum* culture supernatant (11). The assigned ATG initiation codon was preceded by a typical Shine-Dalgarno sequence, AGGAGG, which exhibited perfect complementarity with the 3′ end of *Bacillus subtilis* 16S rRNA (22). Furthermore, the deduced amino acid sequence following the putative ATG resembled the signal sequence preceding *C. thermocellum* endoglucanases A (EGA), B (EGB), and D (EGD), as well as secretory proteins from other gram-positive bacteria (45). Comparison with other cleavage sites (45) suggested that cleavage of the xylanase Z signal peptide may occur between alanine residues 28 and 29. A perfect 14-bp palindrome (Fig. 2), corresponding to an mRNA hairpin loop with a ΔG of −35 kcal/mol (ca. 146 kJ/mol) (38), occurred 30 bp downstream from the TGA stop codon. This structure could act as a transcriptional termination signal (33). No sequence displaying obvious homology with known *E. coli* (33) or *B. subtilis* (19) promoters was found within the 300 bp preceding the start codon (not shown).

Codon usage. Table 1 shows the codon usage of the *xynZ* gene, which was similar to that found for the other sequenced genes of *C. thermocellum*, *celA* (unpublished results), *celB* (10), and *celD* (16). The codon usage of *C. thermocellum* appears to be more closely related to that found in *Bacillus* spp. (40) than in *E. coli* (17).

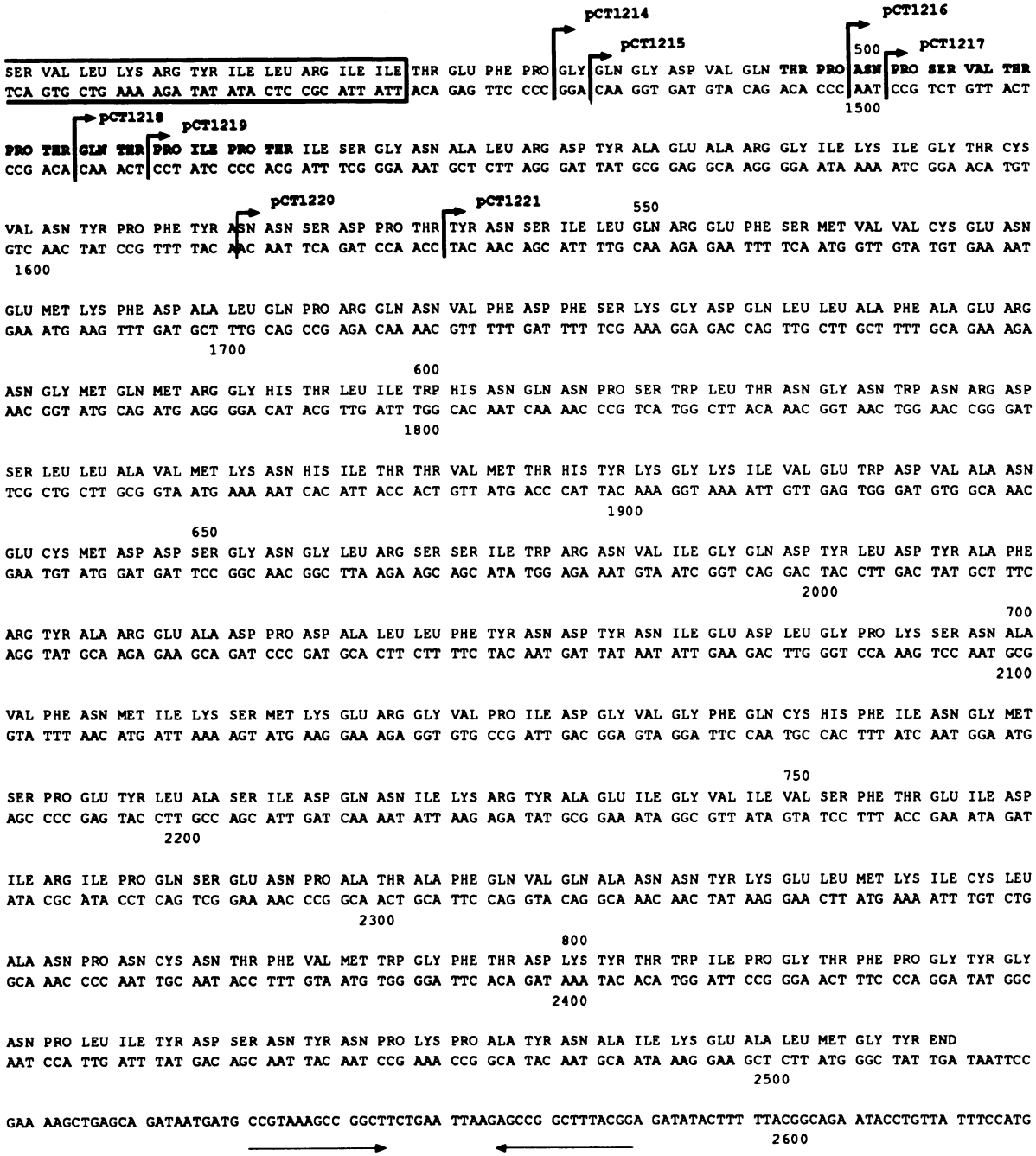
Comparison of xylanase Z with other xylanases and cellulases. No similarity could be found with the xylanases of *Bacillus pumilus* (8, 25) and *B. subtilis* (30). However, a region of 60 amino acids between residues 429 and 488 of xylanase Z was strongly similar to the conserved COOH-terminal region of EGA, EGB, and EGD of *C. thermocellum* (Fig. 3). This region contained two segments of 24 amino acids showing strong sequence similarity, linked by 10 residues. The peptide linking the two segments was different in size and composition in the four proteins, except in EGA and EGB.

Similarities to other regions of *C. thermocellum* cellulases or to cellulases from other microorganisms were much less obvious. However, xylanase Z contained two regions enriched in Pro and Thr residues, located between amino acids 287 and 296 and between amino acids 498 and 514, which were reminiscent of the regions enriched in Pro and Thr found in *C. thermocellum* EGA (2) and EGB (10), *Cellulomonas fimi* endoglucanase (47) and exoglucanase (29), and *Trichoderma reesei* cellulases (31, 36, 37). These regions have been shown to be glycosylated in *T. reesei* cellulases, but the function of these sequences in cellulases from other microorganisms remains to be determined. No convincing similarity was detected with cellulases from *B. subtilis* (20, 26, 32).

Deletion analysis of *xynZ*. The size of the xylanase encoded by *xynZ* was larger than that of most characterized xylanases (7, 8, 13). In order to determine which part of the protein is essential for activity, *xynZ* was sequentially deleted from the 5′ end by using exonuclease III and mung bean nuclease. The remaining part of the gene was religated to the 5′ end of *lacZ* contained in pUC19, yielding plasmids pCT1214 through pCT1221 (Fig. 1 and 2).

The replacement, in clones containing pCT1214 through pCT1219, of the endogenous transcriptional and translational start sites with those controlling the expression of *lacZ* resulted in a drastic increase in xylanase activity (Table 2). The specific activity of TG1(pCT1216) was 220-fold higher than that of TG1(pCT1202), which harbors the original insert. Activity towards pNP-β-D-cellobioside, pNP-β-D-xylopyranoside, pNP-β-D-xylobioside, and MUC followed the same pattern (data not shown). Western blot analysis confirmed that differences of activity were roughly correlated with the amount of antigen present in crude extracts (Fig. 4). The estimated molecular weight of the largest detected polypeptides was somewhat lower than that calculated from the nucleotide sequence. However, the observed

FIG. 2. Nucleotide sequence and deduced amino acid sequence of the *xynZ* gene of *C. thermocellum*. Numbering of both nucleotides and amino acids starts with the beginning of the coding sequence. The putative Shine-Dalgarno sequence (SD) is underlined. Pro- and Thr-rich regions are in boldface type. The conserved, duplicated stretch is boxed (residues 430 to 453 and 464 to 487). A perfect 14-bp palindrome which may serve as a transcription terminator is indicated by inverted arrows. Arrows in the coding sequence indicate the beginning of the *xynZ* gene in the deleted clones.



size variations were generally in good agreement with those expected from the calculated molecular weights. It is not known whether the systematic discrepancy between observed and calculated molecular weights results from proteolytic cleavage occurring at the COOH end of all polypeptides or from inaccuracies of M_r determination by Western blot analysis (39). Proteolysis was most likely responsible for the appearance of lower-M_r species detected in the samples. No strong decrease in specific activity of the xylanase polypeptide was observed with deletions reaching up to 508 residues in the case of pCT1219. The lower specific activity of the pCT1219 crude extract compared with those pCT1214 to pCT1218 appeared to correlate with a lower level of expression of the polypeptide (Fig. 4, lane 7). However, pCT1221, in which 36 further codons were deleted, expressed an immunoreactive polypeptide (Fig. 4, lane 12) which was present in a much larger amount than that encoded by pCT1202, but which had lost all detectable activity.

The xylanase activity expressed from pCT1211 was intermediate between that expressed from pCT1202 and pCT1214 to pCT1219. Since pCT1211 had lost the original *xynZ* start codon and since the remaining part of the gene was fused out of frame with the *lacZ* start codon, this suggests that translation initiation occurred at an internal site within the coding sequence. Deleting up to codon 539, as in pCT1220, abolished reading frame-independent expression of an immunoreactive polypeptide (Fig. 4, lane 11), suggesting that

TABLE 1. Codon usage in the *xynZ* gene of *C. thermocellum*

Amino acid	Codon	No. of codons	Amino acid	Codon	No. of codons	Amino acid	Codon	No. of codons
Phe	UUU	19	Pro	CCU	6	Lys	AAA	23
	UUC	14		CCC	10		AAG	14
				CCA	9			
Leu	UUA	7		CCG	23	Asp	GAU	27
	UUG	19					GAC	20
	CUU	23	Thr	ACU	13			
	CUC	8		ACC	13	Glu	GAA	25
	CUA	0		ACA	18		GAG	11
	CUG	9		ACG	6			
						Cys	UGU	4
Ile	AUU	34	Ala	GCU	9		UGC	5
	AUC	6		GCC	12			
	AUA	22		GCA	20	Trp	TGG	13
				GCG	11			
Met	ATG	19				Arg	CGU	1
			Tyr	UAU	28		CGC	3
Val	GUU	16		UAC	13		CGA	0
	GUC	6					CGG	2
	GUA	19	His	CAU	6		AGA	17
	GUG	5		CAC	5		AGG	11
Ser	UCU	5	Gln	CAA	10	Gly	GGU	21
	UCC	8		CAG	13		GGC	17
	UCA	12					GGA	38
	UCG	11	Asn	AAU	39		GGG	3
	AGU	13		AAC	32			
	AGC	11						

the internal start codon must be located upstream from codon 539. Conversely, although pCT1223 contained a shorter segment of the coding sequence than pCT1211, the polypeptide expressed by pCT1223, in which the gene was fused in frame with *lacZ*, was larger than that expressed by pCT1211. This suggests that translation initiation in pCT1211 must occur downstream from the deletion endpoint of pCT1223, i.e., codon 353. The most likely initiation codon in the region 353 to 539 is GTG (Val-471), which is preceded, 7 bp upstream, by the Shine-Dalgarno-like sequence AGGAG. The length of the corresponding translation product is expected to exceed the translation product of pCT1214 by only 12 residues, and the immunoreactive polypeptides were found to have similar electrophoretic migration (Fig. 4, lanes 3 and 5).

The M_r of the largest immunoreactive polypeptide found in extracts of TG1(pCT1202) was about 90,000, which is compatible with translation initiation occurring at the original start codon. However, a band comigrating with the product expressed from pCT1211 was also clearly detected, suggesting that internal reinitiation occurred within the undeleted gene (Fig. 4, lane 2).

DISCUSSION

The modular pattern found in the sequence of xylanase Z closely parallels the structural organization of several cellu-

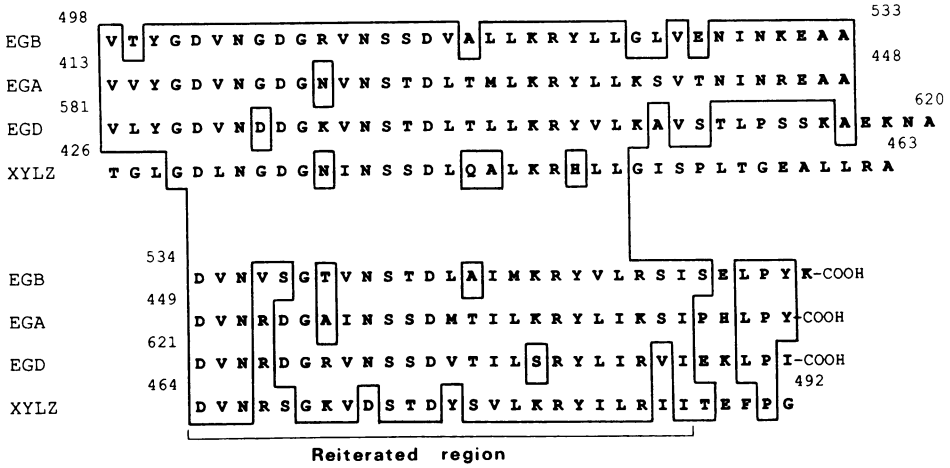


FIG. 3. Alignment of the conserved region between the three endoglucanases EGA, EGB, and EGD and the xylanase Z (XYLZ) of *C. thermocellum*. The eight stretches of 24 amino acids each are aligned. The last displayed amino acid of the three endoglucanases is the COOH-terminal residue of the protein. The boxed amino acids are identical or have similar chemical properties.

TABLE 2. Specific activity of crude extracts of *E. coli* TG1 carrying various subclones of *xynZ*[a]

Plasmid	Sp act on xylan (U/mg)	*xynZ* in frame with *lacZ*	M_r estimated from Western blot	Mol wt calculated from sequence
pCT1202	0.15	−	90,000	92,159
pCT1211	1.7	−	38,500	41,767
pCT1214	24	+	38,500	40,210
pCT1215	30	+	38,500	40,153
pCT1216	33	+	36,500	39,428
pCT1217	33	+	36,500	39,314
pCT1218	23	+	35,000	38,732
pCT1219	8.5	+	35,000	38,503
pCT1220	ND[b]	−	—	—
pCT1221	ND	+	32,000	34,596
pCT1223	26	+	47,000	54,607

[a] Apparent molecular weights were estimated for the largest species detected in the Western blot experiment (Fig. 4). In clones carrying *xynZ* fused in frame with the 5′ end of *lacZ*, the calculated molecular weight includes the few N-terminal amino acids of β-galactosidase which are fused to xylanase Z.
[b] ND, Not detectable; specific activity lower than 0.5% of TG1(pCT1202) specific activity.

lases, in which similar domains are shuffled at different locations within the sequences of various enzymes (37, 44). The phenotype of deletions extending from the 5′ end of *xynZ* indicates that xylanase Z contains a distinct hydrolytic domain located in the COOH-terminal third of the protein. This domain does not include the Pro- and Thr-rich segments or the reiterated region conserved in xylanase Z, EGA, EGB, and EGD. Likewise, the reiterated region can be deleted from *celD* without loss of activity of EGD (unpublished data). Furthermore, another endoglucanase of *C. thermocellum*, EGC, does not possess this repeated sequence (34a). Therefore, this sequence does not belong to the reaction center of xylanase or endoglucanase. However, the strong conservation of this region in the four proteins suggests an important function. One role could be the anchorage of different enzymes to the multimolecular complex (cellulosome) responsible for cellulose hydrolysis. Binding could possibly be mediated by the S1 subunit, a 210-kDa noncellulolytic component which may be involved in the structural organization of the complex and in adhesion to cellulose (1). Alternatively, the reiterated domain could function as a binding site for adjacent subunits of the substrate, similar to the COOH-terminal domain of *T. reesei* cellobiohydrolase I (41). The presence of a similar structure within a xylanase does not preclude either hypothesis, since xylan is structurally similar to cellulose and since xylanase Z is probably associated with the cellulosome (11). No data are available on the function of the N-terminal half of the protein, which could contain another catalytic domain.

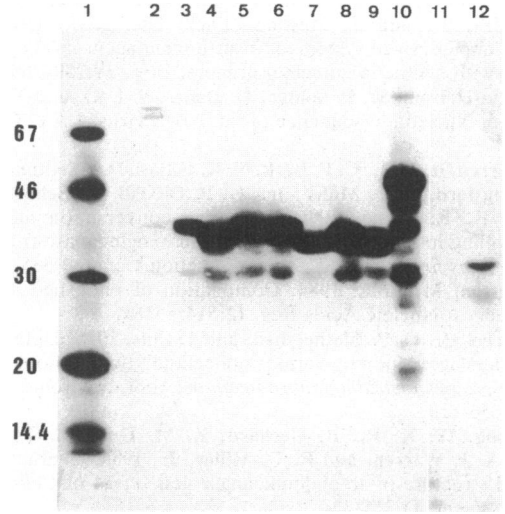


FIG. 4. Detection of the *xynZ* gene product in crude extracts of *E. coli* harboring subclones and deletions of *xynZ*. Samples (60 μg of protein) were denatured for 5 min at 100°C in the presence of 5% β-mercaptoethanol and 2% sodium dodecyl sulfate and analyzed by Western blotting (39). Lane 1, Standard proteins: bovine serum albumin, 67 kDa; ovalbumin, 46 kDa; carbonic anhydrase, 30 kDa; soybean trypsin inhibitor, 20 kDa; lysozyme, 14.4 kDa. Lane 2, pCT1202. Lane 3, pCT1211. Lane 4, pCT1217. Lane 5, pCT1214. Lane 6, pCT1215. Lane 7, pCT1219. Lane 8, pCT1216. Lane 9, pCT1218. Lane 10, pCT1223. Lane 11, pCT1220. Lane 12, pCT1221.

ACKNOWLEDGMENTS

We thank B. Henrissat for pointing out to us the similarity between *C. thermocellum* xylanase Z and *C. fimi* exoglucanase, T. Msadek for his help with the manuscript, and J.-P. Aubert for helpful discussions.

This work was supported by grants ATP 955355 from CNRS and EN3B 0082-F from the Commission of European Communities and by research funds from the University of Paris 7.

ADDENDUM IN PROOF

The segment extending between residues 541 and 833, containing the catalytic site of xylanase Z, shared about 43% identical amino acids with residues 66 to 352 of *C. fimi* exoglucanase (29), which also contain the active site of the enzyme (44). This observation may be correlated with the fact that *C. fimi* exoglucanase also displays strong xylanase activity (N. R. Gilkes, M. L. Langsford, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren, J. Biol. Chem. **259**:10455–10459, 1984).

LITERATURE CITED

1. **Bayer, E. A., E. Setter, and R. Lamed.** 1985. Organization and distribution of the cellulosome in *Clostridium thermocellum*. J. Bacteriol. **163**:552–559.
2. **Béguin, P., P. Cornet, and J.-P. Aubert.** 1985. Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. J. Bacteriol. **162**:102–105.
3. **Biely, P.** 1985. Microbial xylanolytic systems. Trends Biotechnol. **3**:286–290.
4. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
5. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134**: 1141–1156.
6. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**: 165–170.
7. **Dekker, R. F. H., and G. N. Richards.** 1976. Hemicellulases: their occurrence, purification, properties, and mode of action. Adv. Carbohydr. Chem. Biochem. **32**:277–352.
8. **Fukusaki, E., W. Panbangred, A. Shinmyo, and H. Okada.** 1984. The complete nucleotide sequence of the xylanase gene (*xynA*) of *Bacillus pumilus*. FEBS Lett. **171**:197–201.
9. **Garcia-Martinez, D. V., A. Shinmyo, A. Madia, and A. L. Demain.** 1980. Studies on cellulase production by *Clostridium thermocellum*. Appl. Microbiol. Biotechnol. **9**:189–197.
10. **Grépinet, O., and P. Béguin.** 1986. Sequence of the cellulase gene of *Clostridium thermocellum* coding for endoglucanase B.

Nucleic Acids Res. **14**:1791–1799.
11. **Grépinet, O., M.-C. Chebrou, and P. Béguin.** 1988. Purification of *Clostridium thermocellum* xylanase Z expressed in *Escherichia coli* and identification of the corresponding product in the culture medium of *C. thermocellum*. J. Bacteriol. **170**:4576–4581.
12. **Guo, L.-H., and R. Wu.** 1983. Exonuclease III: use for DNA sequence analysis and in specific deletions of nucleotides. Methods Enzymol. **100**:60–96.
13. **Honda, H., T. Kudo, Y. Ikura, and K. Horikoshi.** 1985. Two types of xylanases of alkalophilic *Bacillus* sp. no. C-125. Can. J. Microbiol. **31**:538–542.
14. **Humphreys, G. O., G. A. Willshaw, and E. S. Anderson.** 1975. A simple method for the preparation of large quantities of pure plasmid DNA. Biochim. Biophys. Acta **383**:457–463.
15. **Johnson, E. A., M. Sakajoh, G. Halliwell, A. Madia, and A. L. Demain.** 1982. Saccharification of complex cellulosic substrates by the cellulase system from *Clostridium thermocellum*. Appl. Environ. Microbiol. **43**:1125–1132.
16. **Joliff, G., P. Béguin, and J.-P. Aubert.** 1986. Nucleotide sequence of the cellulase gene *celD* encoding endoglucanase D of *Clostridium thermocellum*. Nucleic Acids Res. **14**:8605–8613.
17. **Konigsberg, W., and G. N. Godson.** 1983. Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80**:687–691.
18. **Lamed, R., E. Setter, and E. A. Bayer.** 1983. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. J. Bacteriol. **156**:828–836.
19. **Losick, R., and J. Pero.** 1981. Cascades of sigma factors. Cell **25**:582–584.
20. **MacKay, R. M., A. Lo, G. Willick, M. Zuker, S. Baird, M. Dove, D. Moranelli, and V. Seligy.** 1986. Structure of a *Bacillus subtilis* endo-β-1,4-glucanase gene. Nucleic Acids Res. **14**:9159–9170.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz.** 1981. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β-lactamase gene. J. Biol. Chem. **256**:11283–11291.
23. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
24. **Millet, J., D. Pétré, P. Béguin, O. Raynaud, and J.-P. Aubert.** 1985. Cloning of ten distinct DNA fragments of *Clostridium thermocellum* coding for cellulases. FEMS Microbiol. Lett. **29**: 145–149.
25. **Moriyama, H., E. Fukusaki, J. Cabrera Crespo, A. Shinmyo, and H. Okada.** 1987. Structure and expression of genes coding for xylan-degrading enzymes of *Bacillus pumilus*. Eur. J. Biochem. **166**:539–545.
26. **Nakamura, A., T. Uozumi, and T. Beppu.** 1987. Nucleotide sequence of a cellulase gene of *Bacillus subtilis*. Eur. J. Biochem. **164**:317–320.
27. **Nelson, N.** 1944. A photometric adaptation of the Somogyi method for the determination of glucose. J. Biol. Chem. **153**: 375–380.
28. **Ng, T. K., and J. G. Zeikus.** 1981. Purification and characterization of an endoglucanase (1,4-β-D-glucan glucanohydrolase) from *Clostridium thermocellum*. Biochem. J. **199**:341–350.
29. **O'Neill, G., S. H. Goh, R. A. J. Warren, D. G. Kilburn, and R. C. Miller, Jr.** 1986. Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. Gene **44**:325–330.
30. **Paice, M. G., R. Bourbonnais, M. Desrochers, L. Jurasek, and M. Yaguchi.** 1986. A xylanase gene from *Bacillus subtilis*: nucleotide sequence and comparison with *B. pumilus* gene. Arch. Microbiol. **144**:201–206.
31. **Penttilä, M., P. Lehtovaara, H. Nevalainen, R. Bhikhabhai, and J. Knowles.** 1986. Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene. Gene **45**:253–263.
32. **Robson, L. M., and G. H. Chambliss.** 1987. Endo-β-1,4-glucanase gene of *Bacillus subtilis* DLG. J. Bacteriol. **169**:2017–2025.
33. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
34a. **Schwarz, W. H., S. Schimming, K. P. Rücknagel, S. Burgschwaiger, G. Kreil, and W. L. Stauden bauer.** 1988. Nucleotide sequence of the *celC* gene encoding endoglucanase C of *Clostridium thermocellum*. Gene **63**:23–30.
35. **Sedmak, J. J., and S. E. Grossberg.** 1977. A rapid, sensitive and versatile assay for protein using Coomassie brilliant blue G-250. Anal. Biochem. **79**:544–552.
36. **Shoemaker, S., V. Schweickart, M. Ladner, D. Gelfand, S. Kwok, K. Myambo, and M. Innis.** 1983. Molecular cloning of exocellobiohydrolase I derived from *Trichoderma reesei* strain L27. Bio/Technology **1**:691–696.
37. **Teeri, T. T., P. Lehtovaara, S. Kauppinen, I. Salovuori, and J. Knowles.** 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. Gene **51**:43–52.
38. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
39. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
40. **Uhlén, M., B. Guss, B. Nilsson, S. Gatenbeck, L. Philipson, and M. Lindberg.** 1984. Complete sequence of the staphylococcal gene encoding protein A. J. Biol. Chem. **259**:1695–1702.
41. **Van Tilbeurgh, H., P. Tomme, M. Claeyssens, R. Bhikhabhai, and G. Pettersson.** 1986. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. FEBS Lett. **204**:223–227.
42. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
43. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alison.** 1985. Nucleotide sequence of the AIDS virus, LAV. Cell **40**:9–17.
44. **Warren, R. A. J., C. F. Beck, N. R. Gilkes, D. G. Kilburn, M. L. Langsford, R. C. Miller, Jr., G. P. O'Neill, M. Scheufens, and W. K. R. Wong.** 1986. Sequence conversation and region shuffling in an endoglucanase and an exoglucanase from *Cellulomonas fimi*. Proteins Struct. Function Genet. **1**:335–341.
45. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5164.
46. **Wiegel, J., C. P. Mothershed, and J. Puls.** 1985. Differences in xylan degradation by various noncellulolytic thermophilic anaerobes and *Clostridium thermocellum*. Appl. Environ. Microbiol. **49**:656–659.
47. **Wong, W. K. R., B. Gerhard, Z. M. Guo, D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1986. Characterization and structure of an endoglucanase gene *cenA* of *Cellulomonas fimi*. Gene **44**:315–324.
48. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.